AUTOMATIC FACIAL EXPRESSION RECOGNITION USING IMAGE PROCESSING AND BAYESIAN REGULARIZED RECURRENT NEURAL NETWORK

Dr. D. Sivakumar 1#1, R. Sofia 2#2

#1 Department of Electronics and Instrumentation Engineering, Annamalai University, Annamalai Nagar, Chidambaram, India.
#2 Department of Electronics and Instrumentation Engineering, Annamalai University, Annamalai Nagar, Chidambaram, India.

ABSTRACT

Facial expressions can be considered as a means of communication by non-verbal signals. They are essential part of human relations. Automatic facial expressions recognition can be imperative for natural human-machine interaction. Automatic facial expressions recognition can be utilized in the field of behavioral science and in the health care department. Although humans perceive the facial expressions immediately and effortlessly, reliable automatic facial expression recognition by a machine is a challenging task. This paper proposes a hybrid method, comprising of image processing and Artificial Neural Networks (ANN) for automatic facial expression recognition. Image processing of the faces were carried out to extract the features. The facial features include the change in the shape and size of eyebrows, eyes and lips. The features extracted from the image are quantified and are used as training set for the artificial neural network. Two types of ANNs viz. feedforward neural network and Bayesian regularized recurrent neural network. Six types of expressions are recognized. The results are compared on the basis of confusion matrix, error histogram, mean absolute error, error plot and regression plot. On all the evaluation parameters Bayesian regularized recurrent neural network (BRRNN) is found to be best suited for automatic facial expression recognition.

Key words: Facial expression recognition, Bayesian regularized recurrent neural network, Feedforward neural network, Gamma Correction, Feature extraction, Image enhancement.

Corresponding Author: Dr. D. Sivakumar, R. Sofia

INTRODUCTION

Emotion is a mental state that happens spontaneously rather than through conscious attempt and it is accompanied by physical changes. Human computer interaction can be automated by using emotions and the accompanied physiological changes. The computers not only respond to human commands, they also try to mimic the human behavior. Automatic recognition of human
Facial expression and emotion takes the computer one step near the above said task. Human emotion recognition is an interesting and challenging problem. As the proverb ‘Face is the index of the mind’ rightly suggests, facial expression reflects the feeling and emotions felt by a person.

Whenever a user watches video or listens to music, his facial expressions changes. In order to translate facial expressions to emotions, emotion assessment technique are required. Emotion assessment is a difficult task since users are not always able to express their emotion with words all the time and the self-reporting emotions have a high probability of false emotions. In this work, a real-time automatic facial expression recognition system using image is proposed.

For efficient detection of spatial expression, many algorithms have been proposed. Ghimireet al[2] proposed extraction of local region specific features and support vector machines for classification of the facial expression. Classification accuracy of 97% was achieved. Shan et al[3] applied support vector machine classifiers on the boosted local binary pattern features obtained from different image data sets. The accuracy obtained for jaffe image data set was 81%. Cohn-Kanade database of video signal is classified by Siddiqi et al[4] using Hidden Conditional Random Fields and a classification accuracy of 93% is achieved. Happy and Routray[4] implemented facial expression recognition from jaffe and ck images by determining facial patches.

Zhonget al[1] also determined facial patches and used them for multitask sparse learning framework. But facial patches vary for different sets of faces. Kotsia and Pitas[6] used geometric deformation features and support vector machines for facial expression detection of Cohn-Kanade database and obtained a classification accuracy of 99.7%. Huang and Tai[7] used key point descriptors for determining the features and classified the expression using weighted majority voting classifier. The accuracy obtained for jaffe image data set was 93.33%. Chiet al[8] suggested a cloud model of images for expression recognition. The effectiveness of the suggested method is not mentioned. Sarode and Bhatia[9] extracted intransient facial features and detected four facial expressions. An average accuracy of 81.5% is obtained for four classes of jaffe image data set. Shruti Bansal and Pravin Nagar[10] used Bezier curves for determining emotions. The accuracy obtained was 70%.

Numerous attempts were made for effective recognition of emotion from the facial images. Mostly jaffe image dataset or Cohn-Kanade database were used for carrying out the research. From the literatures it has been found that the maximum accuracy obtained for jaffe data set is 93.33% which is lesser than that of Cohn-Kanade database. Hence for our work, jaffe image data set is chosen. The geometrical features are extracted from images and ANN is used as classifier for emotion detection.

Kezhenglin, Weiyue Cheng and Jingtian Li et al[12] attended with Geometric features collection and calculated 3D space of false Geodesic distance for FER (Facial Expression Detection). Jagdish Lal Raheja and Umesh Kumar et al[13] experimented a Viola and Jones using Add boost Haar classifier for face detection and then edge detection, thinning and token detection are performed of FER. Xiaoming Chen and Wushan Cheng et al[14], paper uses a canny edge detection method for facial expression recognition. Deepak Ghimire,
Sunghwan Jeong, Joonwhoan Lee, Sang Hyun Park et al [15], have used a Regions specific appearance features, geometric features and the important local region are determined for improvement in FER accuracy.

Usha Mary Sharma, Jayanta Kumar Das, Trinayan Dutta et al [16], shows that FER can be achieved upto 95% accuracy with jaffe database compared to other database. Aleix Martinez, Shichuan Du et al [17] have presented a paper stating that, both continuous and categorical model have problem in explaining combination of emotion, it is resolved by a model consist of C distance continuous spaces. Mikael Bod’en et al [18] provides guidance to some of the concepts surrounding recurrent neural network. Kim Hartmann, Ingo Siegert, Stefan Gluge et al [19] used mathematical modeling as a unifying language to translate the coherence of appraisal theory. Phd. Gheorghe Gilca, Professor Nicu-George Bizdoaca et al [20] used a Viola Jones algorithm detects human face and then it is transferred to sugeno type decisional fuzzy system, and fuzzification measurements of eyes, eyelids and mouth is done for FER. V. Gomathi, Dr. K. Ramar and A. Santhiyaku Jeerkumara et al [21] shows the facial expressions are recognized using multiple adaptive neuro fuzzy inference system (MANFIS). The Proposed system tested with JAFFE face database. Simina Emerich, Eugen Lupu, Anca Apati et al [22] used a bimodal emotion recognition system using the combination of facial expressions and speech signals.

Deepthi S, Archana G.S, Dr. Jagathy Raj V.P et al [23] paper proposes a method using artificial neural networks to find the facial expressions among the three basic expressions given using MATLAB (Neural Network) toolbox. Ahmad R. Naghsh-Nilchi and Mohammad Roshanazamir et al. [24] this paper uses a optical flow approach which is examined over cohn-kanade AU-Coded Facial Expression Database, which could identify facial expression correctly up to 94%. Li Xia et al. [25] shows that multiclassification method based on SVM can obviously reduce the training and testing time and improve the classification performance. M. Singh, A. Majumder, L. Behera et al. [26] have presented a facial expression recognition system using Bayesian network.

The rest of the paper is organized as follows. The feature extraction from facial images is described in section II. Section III presents a detailed insight into the artificial neural networks applied for facial expression recognition. In Section IV the performance of automatic facial expression recognition system is analyzed. In Section VI concluding remarks are given. The facial expression recognition has two stages viz. the feature extraction and the expression recognition. The feature extraction is carried out by image processing.

**FEATURE EXTRACTION**

The flow diagram for feature extraction by image processing is shown in Figure 1. Initially the human face features are detected from the original image by removing the neck and part of hair. This is achieved by cropping the image. The cropped image is fed as input to image enhancement unit. In this, contrast adjustment is carried out. Gamma correction is performed with a gamma value of 1.25. Gamma value greater than 1, expands the lighter regions and compresses the darker regions of the original image. This makes the facial features of the
images to stand out. For removal of hair a mask is created from the preprocessed image by converting the image into black and white and removing smaller structures.

The enhanced image is obtained from the preprocessed image by converting the image into black and white and complementing the result. By ANDing the mask with the enhanced image facial features are obtained by their orientation and size. The facial features extracted are eyes, mouth and eyebrows. For the extracted features the geometrical feature values viz. area, orientation, perimeter, solidity, major axis length, minor axis length and centroid are determined. As centroid has x and y values, totally 40 features (eight geometrical features of five facial features (two eyes, two eyebrows and mouth)) are obtained for each image.

![Diagram](image.png)

**Fig 1: Feature Extraction by Image Processing**

(a) Original Image  
(b) Cropped Image  
(c) Preprocessed Image
The outputs of feature extraction steps are shown in Figure 2. Then the facial features are classified into different expressions. The facial features include the change in the shape and size of eyebrows, eyes and lips. The features extracted from the image are quantified and are used as training set for the neural network.

**ARTIFICIAL NEURAL NETWORK AS CLASSIFIER**

An artificial neural network (ANN), usually called as neural network (NN), is a mathematical model or computational model that simulates the structure and/or functions of biological neural networks (Haykin 1999)[11]. It consists of a network of interconnected artificial neurons which processes information and generates the relevant result. A neural network can be characterized by the following points.

1. It is composed of essential elements called neurons. These neurons operate in parallel.
2. Connections between the neurons and the weights associated with them determines the neural network function
3. The input signals arrive at the neurons by traveling through these connection links.
4. The signal passing through these connection links are multiplied by the link weights.
5. Each neuron has a transfer function known as activation function.
6. The output of the neuron is determined by the activation function.

There are many types of neural networks available. Perception Network, feedforward network, recurrent network, Back propagation networks, Self Organizing Maps are some of the commonly used architectures.

There are two types of classifications. They are supervised and unsupervised classification. As expressions are unique for each and every individual, the features are extracted are given as training set for the neural network. The facial expression is detected by supervised classification. The trained neural network in future will classify the images according to the emotions expressed by the person.
Feedforward Neural Network (FFNN)

A feedforward neural network is a biologically inspired classification network. In this, information is passed in the forward direction only (Simon Haykin 1999). In a feedforward network neurons are formed in layers, where the first layer takes inputs and the last layer delivers the outputs. The middle layers which have no association with the outside world are called hidden layers. Each neuron in one layer is connected to every neuron in the next layer. As information is constantly "fed forward" from one layer to the next these networks are called feed forward networks. There is no connection between neurons in the same layer.

The hidden layer neurons of a Feed Forward Neural Network take sigmoid activation function. The sigmoid activation function is given by,
Linear activation function for the output layer has been used. These neurons give the output directly proportional to their input. i.e. the input and output have linear relationship. Back propagation learning algorithm is used for training.

Figure 3 shows the Feed Forward Neural Network. The network is assumed to have n inputs denoted by $x_i$ where $i$ varies from 1 to n, m outputs denoted by $y_k$ where $k$ varies from 1 to m. There are n input neurons and m output neurons. The hidden layer has p neurons. The connection between layers and their weights are shown in figure. The weights of the input to hidden layer connections are marked as $v$ and the weights of hidden layer to output layer links are marked as $w$.

**Bayesian regularized recurrent neural network**

Recurrent Neural Networks are fundamentally different from feed forward architectures. A Recurrent Neural Network (RNN) is a modification to this architecture to allow for temporal classification (Jordan 1986). In this there are closed loop paths from a unit back to itself. i.e. information is sent from later stages to prior stages. A `context" layer is augmented to the network, which holds information between steps. At every time instant, new sources of information are supplied into the RNN. The prior contents of the hidden layer are passed into the context layer. These are then supplied once again to the hidden layer in upcoming steps. In this way, the network keeps a short term memory. As there is feedback, training is fast and accurate.

Elman network is a recurrent network which uses back propagation algorithm. It is invented by Jeff Elman (1990). It has two layers namely, one output layer and one hidden layer. There is a feedback connection from the output of the hidden layer to its input. Hence, the hidden layer is also known as recurrent layer. This feedback path helps the network to learn to recognize and generate temporal as well as spatial patterns.

Recurrent network hidden layer neurons are hyperbolic tangent sigmoid neurons. i.e. the activation function or transfer function of the neurons in this layer is hyperbolic tangent sigmoid function. The hyperbolic tangent sigmoid activation function is given by,

$$F(n) = \frac{1}{1 + \exp(-n)}$$  \hspace{1cm} (3.1)

This is mathematically equivalent to $\tanh(n)$. The maximum output limits are -1 and 1.

Recurrent network has linear neurons in its output layer. These neurons give the output directly proportional to their input. The hidden layer of this network must have enough number of neurons to get the correct output.

$$F(n) = \frac{2}{1 + \exp(-2n)} - 1$$  \hspace{1cm} (3.2)
This network differs from the conventional two layer network, since it has feedback or recurrent connection in its first layer. The delay in the feedback connection stores the previous values which can be used in the current instant. Even if two Elman networks of same weights and biases are fed, the identical inputs, at the same instant, their outputs can be different, due to different feedback states.

Fig 4: Bayesian Regularized Recurrent Neural Network Architecture
Figure 4 shows the Recurrent Neural network. The network is assumed to have n inputs denoted by $x_i$ where $i$ vary from 1 to n, m outputs denoted by $y_k$ where $k$ varies from 1 to m. There are n input neurons and m output neurons. The hidden layer has p neurons. The connection between layers and their weights are shown in Figure. The connection shown by the bold lines indicates the delayed output fed back to the input. D block indicates the delay. The weights of the feedback connection are denoted by $u$. The weights of the input to hidden layer connections are marked as $v$ and the weights of hidden layer to output layer links are marked as $w$. Elman network has the advantage that it can store information for future reference. It can be trained to respond to spatial and temporal patterns.

The training mechanism for the Elman network (Elman 1993) can be summarized in the following steps.

1. Initialize the activations of the context neurons to zero.
2. Determine the input vector of the hidden layer, by concatenating the external input $(x(t),...,x(t-D))$ at instant $t$ and the neurons context activations at instant $t$.
3. Propagate the input to output for obtaining the prediction at instant $t+1$.
4. Determine the weights of the network by Bayesian regularized back propagation algorithm.
5. Increase the time variable by one unit and go to step 2.

Once a neural network is trained with an input output pattern, then it is ready for real time application. The test data is then fed to the network and desired output is obtained. A properly trained neural network tends to give reasonable answers when the inputs are different from the training data. New output for unknown input is similar to the correct output of trained input which is similar to the unknown input. This property of the neural network makes it possible to train a network with a set of input output pattern and get good results without training it with all possible combinations of input output pairs.

The training algorithm used for recurrent neural network is Bayesian regularized back propagation algorithm. The steps of the training algorithm are given below.

1. Determine the Jacobian matrix $J$. It is a matrix of first order derivatives of output vector with respect to the input vector.
2. Obtain the error gradient $g = J^tE$. Where $E$ is the error vector.
3. Determine the Hessian matrix which is the cross product of Jacobian matrix
   \[ H = J^tJ \]
   Obtain the cost function
   \[ C = \beta E_{se} + \alpha E_{sw}, \]
   where:
   - $E_{se}$ is the sum of squared errors, and
   - $E_{sw}$ is the sum of squared weights
   - $\alpha, \beta$, and $\gamma$ are the Bayesian hyper parameters
   Solve $(H + \lambda I)\delta = g$ to obtain $\delta$
   Update the network weights $w$ using $\delta$
   Determine the cost function using the updated weights
   If the cost has not decreased,
Discard the new weights, increase $\lambda$ using $v$ and go to step 5.
Else decrease $\lambda$ (damping factor) using $v$ (adjusting factor)
Update the Bayesian hyper parameters
\[
\gamma = W - (\alpha * \text{tr}(H^{-1}))
\]
\[
\beta = (N - \gamma) / 2.0 * E_{se}
\]
\[
\alpha = W / (2.0 * E_{sw} + \text{tr}(H^{-1})) \text{ or }
\alpha = \gamma / (2.0 * E_{sw})
\]
$W$ is the number of weights and biases
$N$ is the number of entries in the training set
$\text{tr}(H^{-1})$ is the trace of the inverse Hessian matrix
The geometrical features extracted for different classes of expression are given to the neural network as training input – target class pair. The test inputs are given to the trained neural network. The detected expression is shown as a text message.

**PERFORMANCE EVALUATION OF AUTOMATIC DETECTION SYSTEM**

An automatic facial expression recognition system has been developed using image processing and artificial neural networks. The system performance is tested with the jaffe image dataset. The performance is evaluated using the evaluation parameters viz. confusion matrix, error histogram, mean absolute error, error plot and regression plot.

**Error histogram**

The dispersion of the system errors is depicted by the error histogram plot. This histogram is a measure of anomalies. The anomalies are the data points where fit between the original class and the target class significantly worse than the majority of data.

![Error Histogram with 20 Bins](image1)

![Error Histogram with 20 Bins](image2)

**Fig 5: error histogram of FFNN system**

The error histogram of FFNN system is shown in Figure 5. In this case of SVM, it can be seen that the maximum errors fall between -0.05 and 0.1. The error span is from -4 to 2 which shows that some of the expressions are wrongly recognized.
The error histogram of FFNN system is shown in Figure. In this case of SVM, it can be seen that the maximum errors fall between -0.1 and -0.03. The error for entire input falls within this span which indicates that all the expressions are correctly recognized.

**Confusion Matrix**

The performance of a classifier is better evaluated by the confusion matrix. It is an error matrix portrayed by a table layout which helps in visualizing the performance of classifier. The columns of the confusion matrix represent the instances in the output class whereas the rows represent the instances in the actual class.

Figure 6 shows the confusion matrix of FFNN system. It is seen that 83.33% of class A are correctly classified as class A whereas 16.67% are wrongly classified as belonging to class B, C and E. Class C data is correctly classified. Similarly, other expressions are also wrongly classified. Only the last expression is properly identified.

Figure shows the confusion matrix of BRRNN system. The diagonal elements of this matrix are 1 and off-diagonal elements are 0. This shows that, there is exact identification of the expression by BRRNN system.

**Regression Plot**

The regression plot is drawn between the system outputs and actual classes. Figure 7 shows the regression between them. The R value of 1 indicates perfect fit.

In FFNN system during training, the regression curve almost fits which is indicated by the R value 0.99995. In case of validation, the regression is more. While testing it is somewhat improved to the R value of 0.77144. The regression plot of entire data indicates an acceptable regression value of 0.84409.
In BRRNN system the R values of regression plot of training, validation, testing and for entire data are equal to 1. It results in perfect fit. The system outputs and actual expressions are one and the same.

![Regression Plot]

**Fig 7: Regression Plot**

**Error plot**

Error plot is the plot between error and the input instant. Figure 8 shows the error at each instant of time. For FFNN system the error plot is shown in Figure. The error varies from -4 to 2. The horizontal line indicates zero error. The error plot of BRRNN system is shown in Figure. It is a horizontal line at zero error. It indicates no error for all the inputs.

![Error Plot]

**Fig 8: Error Plot**

**Mean Absolute Error (MAE)**

Mean absolute error is a measure of error between the output class of data points and the actual class of the same. It is represented by
Where $y(i)$ is the output class of the $i^{th}$ data point and $t(i)$ is the target class of the same.

For feedforward network the mean absolute error is found to be 29.17%, whereas Bayesian regularized recurrent neural network resulted 0 MAE. This shows that in BRRNN system, there is absolute match of output class and actual class. This indicates that BRRNN system is 100% effective in recognizing the facial expression.

**GRAPHICAL USER INTERFACE**

A graphical user interface has been developed for automatically recognizing the facial expression. The facial image is given as input to the expression recognition system. The features extracted are shown in figure 9. From the features the type of expression recognized is displayed for two both FFNN system and BRRNN System Figure shows the GUI. The input image of face registering fear is wrongly classified by FFNN as angry. The emotion is properly identified by BRRNN System as ‘Afraid’.

![Fig 9: Facial Expression Recognition](image)

**CONCLUSION**

Various endeavors were made for determination of emotion from the facial pictures. The objective of this research is to develop a robust facial expression recognition system. Earlier work in this field has been carried out by identification of nodes in the face or by means of facial patches. A novel feature extraction algorithm based on image processing is applied. The algorithm created a mask from the enhanced image and the mask is utilized for extracting facial features. The facial features are then converted to geometric values and used for training the neural networks. Feedforward as well as feedback neural network are used for expression recognition. The neural network with feedback chosen for the classification is Bayesian regularized recurrent neural network. The performance of the system developed is tested with the jaffe image dataset. The expressions chosen are happy, angry, normal, sad, afraid and disgusted. The expressions are perfectly identified by BRRNN. The FFNN system showed a mean absolute error of 29.17%, The BRRNN system resulted in 0% error. The accuracy of
BRRNN system is 100%. The BRRNN system is 100% effective in recognizing facial expressions. Though this result is remarkable, the future work will involve the expression identification from real time images. The classification with the help of neuro fuzzy system can be tested for real time images. Furthermore, the expression identification can be converted to trauma identification and a useful clinical application can be created.

REFERENCES


[16] International journal of science, Technology and management volume No. 04, Special Issue No. 01, March 2015 ISSN(online): 2394-1537, PERFORMANCE ANALYSIS OF THE TECHNIQUES EXPLOYED ON VARIOUS DATASETS IN IDENTIFYING THE HUMAN FACIAL EMOTION, Usha Mary Sharma, Jayanta Kumar Das, Trinayan Dutta


